Algorithms and Probability

Exercise Session 7

https://n.ethz.ch/~ahmala/anw

Aufgabe 1 – *Zufällige Schnitte*

(a) Für einen Graphen $G = (V, E)$ mit n Knoten und m Kanten betrachten wir den Laplace-Raum $\Omega = \{S \mid S \subseteq V\}$ und die Zufallsvariable $X :=$“Anzahl Kanten über den Schnitt $(S, V \setminus S)$”.

Berechnen Sie $\mathbb{E}[X]$.

Hinweis: Schreiben Sie X als Summe von geeigneten Indikator Zufallsvariablen.

(b) Folgern Sie aus Ihrem Ergebnis in (a), dass G einen Schnitt der Grösse mindestens $m/2$ hat.

Lemma 2.29. Ist X eine Zufallsvariable, so gilt:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega].$$

(b) Folgern Sie aus Ihrem Ergebnis in (a), dass G einen Schnitt der Grösse mindestens $m/2$ hat.

When is X independent of itself?

$E[XY] = E[X] * E[Y]$ <= X and Y are independent

$VAR[X + Y] = VAR[X] + VAR[Y] \quad \Leftarrow \quad$ X and Y are independent

VAR[XY] = VAR[X] * VAR[Y] <?=?> X and Y are independent

$$VAR[X + Y] = VAR[X - Y]$$

VAR[X] = VAR[- X]

Problem

You are trying to cross the street around Zurich HB. On average, one car goes by (i.e. passes along right in front of you) every 2 seconds. Assume that the cars behave independently; also, their directions of travel don't matter, and they go by instantaneously.

a) What is the probability that you will see exactly 1 car in the interval between 1 and 4 seconds from now?
b) You know it takes you 6 seconds to cross the street, and you can't stop in the middle, so you need to wait for an interval of at least 6 seconds between successive cars before you start to cross. (You can see the cars coming, so you know when it is safe to go.) What is the probability that you can start crossing right now?

Problem

$X \sim \mathrm{Po}(\lambda)$

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{for } x \in \mathbb{N}_0 \\ 0 & \text{else} \end{cases}$$

$\mathbb{E}[X] = \lambda$ $\mathrm{Var}[X] = \lambda$

You are trying to cross the street around Zurich HB. On average, one car goes by (i.e. passes along right in front of you) every 2 seconds. Assume that the cars behave independently; also, their directions of travel don't matter, and they go by instantaneously.

a) What is the probability that you will see exactly 1 car in the interval between 1 and 4 seconds from now?
b) You know it takes you 6 seconds to cross the street, and you can't stop in the middle, so you need to wait for an interval of at least 6 seconds between successive cars before you start to cross. (You can see the cars coming, so you know when it is safe to go.) What is the probability that you can start crossing right now?

Satz 2.67. *(Ungleichung von Markov)* Sei X eine Zufallsvariable, die nur nicht-negative Werte annimmt. Dann gilt für alle $t \in \mathbb{R}$ mit $t > 0$, dass

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Oder äquivalent dazu $\Pr[X \geq t \cdot \mathbb{E}[X]] \leq 1/t$.

Satz 2.68. *(Ungleichung von Chebyshev)* Sei X eine Zufallsvariable und $t \in \mathbb{R}$ mit $t > 0$. Dann gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}$$

oder äquivalent dazu $\Pr[|X - \mathbb{E}[X]| \geq t\sqrt{\mathrm{Var}[X]}] \;\leq\; 1/t^2$.

Satz 2.70 (Chernoff-Schranken). Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Dann gilt für $X := \sum_{i=1}^{n} X_i$:

(i) $\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2 \mathbb{E}[X]}$ für alle $0 < \delta \leq 1$,

(ii) $\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{2}\delta^2 \mathbb{E}[X]}$ für alle $0 < \delta \leq 1$,

(iii) $\Pr[X \geq t] \leq 2^{-t}$ für $t \geq 2e\mathbb{E}[X]$.

Consider a test with 18 questions and two possible answers to each question. If one guesses each answer, what is the probability of answering correctly to at least 5 and at most 13 answers?

Give bounds for the probability using the bounds.

Satz 2.67. *(Ungleichung von Markov)* Sei X eine Zufallsvariable, die nur nicht-negative Werte annimmt. Dann gilt für alle $t \in \mathbb{R}$ mit $t > 0$, dass

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Oder äquivalent dazu $\Pr[X \geq t \cdot \mathbb{E}[X]] \leq 1/t$.

Satz 2.68. *(Ungleichung von Chebyshev)* Sei X eine Zufallsvariable und $t \in \mathbb{R}$ mit $t > 0$. Dann gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$$

oder äquivalent dazu $\Pr[|X - \mathbb{E}[X]| \geq t\sqrt{\mathrm{Var}[X]}] \leq 1/t^2$.

Satz 2.70 (Chernoff-Schranken). Seien $X_1, \dots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Dann gilt für $X := \sum_{i=1}^{n} X_i$:

(i) $\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2 \mathbb{E}[X]}$ für alle $0 < \delta \leq 1$,

(ii) $\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{2}\delta^2 \mathbb{E}[X]}$ für alle $0 < \delta \leq 1$,

(iii) $\Pr[X \geq t] \leq 2^{-t}$ für $t \geq 2e\mathbb{E}[X]$.

Satz 2.70 (Chernoff-Schranken). Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Dann gilt für $X := \sum_{i=1}^n X_i$:

(i) $\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2 \mathbb{E}[X]}$ für alle $0 < \delta \leq 1$,

(ii) $\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{2}\delta^2 \mathbb{E}[X]}$ für alle $0 < \delta \leq 1$,

(iii) $\Pr[X \geq t] \leq 2^{-t}$ für $t \geq 2e\mathbb{E}[X]$.

A factory produces 45'000 screw per day. The screws are supposed to be 30 mm long, but sometimes manufacturing errors add up, and some screws have to be discarded. The probability that any given screw is too short is 1/3 and the probability that a screw is too long is 1/5 (the different screws' lengths are independent).

1. Let X be the number of screws discarded over a whole day. How is X distributed? Calculate the expected value μ.

2. Tom is a supervisor at the factory and if $0.9\mu < X$ and $X < 1.2\mu$, he reports to his boss that the day was uneventful. Use Chernoff to give an upper bound on the probability that a given day is eventful.

3. Another factory produces nails and its supervisor tells Tom that their expected number of discarded nails per day is not more than 50. Without knowing anything else about the other factory (except that their nails are also produced independently and have identical chances of being too long or too short), use Chernoff to give an upper bound on the probability that at least 300 nails have to be discarded in a given day.

1. $X \sim \text{Bin}(45'000, \frac{1}{3} + \frac{1}{5} = \frac{8}{15})$. Thus the expected value is $\mathbb{E}[X] = np = 45'000 \cdot \frac{8}{15} = 24'000$.
2. We use Chernoff to get bounds for $\Pr[X \geq 1.2\mathbb{E}[X]]$ and $\Pr[X \leq 0.9\mathbb{E}[X]]$:

$$\begin{aligned} &\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2\mathbb{E}[X]} \\ \implies &\Pr[X \geq 1.2\mathbb{E}[X]] \leq e^{-\frac{1}{3}\cdot 0.2^2 \cdot 24000} = e^{-320} \end{aligned}$$

$$\begin{aligned} &\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{2}\delta^2\mathbb{E}[X]} \\ \implies &\Pr[X \leq 0.9\mathbb{E}[X]] \leq e^{-\frac{1}{2}\cdot 0.1^2 \cdot 24000} = e^{-120} \end{aligned}$$

And we get:

$$\begin{aligned} &\Pr\left[\overline{0.9\mu < X < 1.2\mu}\right] = \Pr[X \geq 1.2\mathbb{E}[X]] + \Pr[X \leq 0.9\mathbb{E}[X]] \\ \implies &\Pr\left[\overline{0.9\mu < X < 1.2\mu}\right] \leq e^{-320} + e^{-120} \approx 7.67 \cdot 10^{-53} \end{aligned}$$

3. Using the normal inequality is not possible because we'd need to choose a $\delta > 1$.
Using the third inequality instead (which we're allowed to do because $300 = 6 \cdot 50 \geq 6 \cdot \mathbb{E}[X] \geq 2e \cdot \mathbb{E}[X]$) gives:

$$\Pr[X \geq 300] \leq 2^{-300} \approx 4.9 \cdot 10^{-91}$$

Las Vegas Algorithm

Always outputs the correct answer

Running time is random

Monte Carlo Algorithm

Always run in a fixed amount of time

Result may be incorrect

In Class Probability Task

https://n.ethz.ch/~ahmala/anw/material/Probability_Task